RR 94-04

MODELS FOR ASYMMETRIC PROXIMITIES

Berrie Zielman
Willem J. Heiser

Department of Data Theory
University of Leiden

Abstract

Models for asymmetric proximities are a combination of a symmetric similarity component and an asymmetric dominance component. The differences and similarities between the methods that are discussed in this paper are revealed by applying a certain decomposition to the model parameters, clearly separating the dominance and symmetric similarity component. The notion of skew-symmetry turns out to be an often seen element in modeling asymmetry, although sometimes in disguise and difficult to recognize.

The decomposition shows that there are two classes of models, one that assumes that the asymmetric relationships are transitive, the other class consists of models that can also represent circular asymmetric relationships. Using this decomposition a classification into two categories is proposed.

Key words: multidimensional scaling, asymmetry, skew-symmetry, similarity-bias model, DEDICOM, unfolding, feature matching model, distance density model, choice model, additive tree

Introduction

Tables where the rows and columns classify the same set of stimuli occur in several research situations. For instance in an identification experiment, where subjects are instructed to identify stimuli, the rows correspond to the presented stimulus and the columns correspond to the stimuli given as responses. Such a table is square, and on the diagonal we find the number of correct responses, the off-diagonal elements correspond to errors or confusions, which may tell us something about the similarity of the stimuli. We often observe asymmetry in such tables, that is, stimulus i is more often confused with stimulus j than the other way round. Sociometric ratings provide another example of an asymmetric data table, where rows refer to children as judges and the columns refer to the same children but now they are being judged. Again asymmetry may be observed because child i could express a higher like or dislike rating to child j than vice versa (Collins, 1987). Other examples include brand switching counts (Rao & Sabavalla, 1979), counts of telephone calls among cities (Harshman & Lundy, 1990), citations among journals (Coombs, 1964; Weeks & Bentler, 1982), first choice-second choice connections (Urban, Johnson & Hauser, 1984), communication and volume flows (Chino, 1978), migration-rates (Tobler, 1976), occupational mobility tables (Clogg, 1981), same-different errors (Rothkopf, 1957), and word associations (Tversky & Hutchinson, 1986).

Another type of data which may yield asymmetric similarity measures are *conjoint distances* (Coombs, 1964, p. 44). Conjoint distances occur when pairs of pairs of stimuli are compared and both pairs have a stimulus in common. Examples of observational schemes for these data are: picking k out of n-1 stimuli that are most similar to the reference item, ordering k out of n-1 objects in terms of relative similarity to the reference item, the method of n-dimensional rank-order, the method of anchor-point ordering and the method of triads.

Asymmetry in *similarity ratings* has been demonstrated by Tversky (1977), who found that stimuli that are prototypes of a stimulus domain are judged less similar to their variants than the other way round. In this study, North Korea was judged more similar to China than China to North Korea. Rosch (1975) observed that the perceived distance from a prototype or focal element to a variant was greater than the distance from the variant to the prototype.

Nosofsky (1991) related other properties of stimuli such as salience, hierarchical status, good stimulus, easily encoded stimulus and high-frequency stimulus to the asymmetry in confusions between stimuli. Similarly, large cities interact more often with small cities than the reverse (Tobler, 1976; Constantine & Gower, 1982); asymmetry in brand switching may indicate the properties of brands that contribute to the success of attracting consumers from brands that do not share these characteristics (DeSarbo & De Soete, 1984); in sociometric research, asymmetry in liking can be a measure of the popularity of a child (Collins, 1987).

In this paper, it is shown that all major asymmetric models in the literature can be decomposed into a symmetric part and a "pure" asymmetric part, called skew-symmetry, by applying a decomposition theorem from linear algebra. The decomposition is used in the literature (Gower, 1977) to decompose the *data* into a symmetric and skew-symmetric part, where these two components are analyzed separately. A psychological interpretation of this decomposition of the data is that the symmetric part can be viewed as a similarity part, whereas the skew-symmetric part can be interpreted as a dominance or preference component. In the present paper, the decomposition is applied not only to data matrices, but also to *subsets of model parameters*. This provides a separation of the *model* into a symmetric component and a skew-symmetric component. At various places throughout this paper, we replace old parameters by new ones that are more easily interpreted in terms of similarity and preference than the original parameters.

We start our discussion with the decomposition of an asymmetric matrix, and we discuss some related results. In the next section we apply this decomposition to the similarity bias model, and we give an overview of its special cases. It is shown that the similarity bias models and its special cases assumes asymmetry as a transitive relation. Because of this transitive relationship, we may also say that these models relate the asymmetry between stimuli to properties of the individual stimuli. We continue with a section where we discuss models that go beyond the similarity bias model, that is, additional asymmetry is explained by a pairwise relation between two stimuli. This class of models can represent circular or intransitive asymmetric relationships. We conclude our exposition with a summary and discussion.

Additive Decomposition of Asymmetry

Any square non-symmetric matrix **Q** with n rows and n columns can be additively decomposed into a symmetric and a skew-symmetric matrix,

$$\mathbf{Q} = \mathbf{S} + \mathbf{A}, \quad (1)$$

where **S** is a symmetric matrix of averages $s_{ij} = \{q_{ij} + q_{ji}\}/2$, with ($i$=1,...,$n$; j=1,...,n), and **A** a skew-symmetric matrix with elements $a_{ij} = \{q_{ij} - q_{ji}\}/2$. This result can be found in a number of textbooks on linear algebra, see for instance Noble and Daniel (1988, p. 20). The property $a_{ij} = -a_{ji}$ is called *skew-symmetry* and sometimes *anti-symmetry*. The matrix **A** describes the departures from symmetry, and can be viewed as the preference or dominance part of an asymmetric matrix; if $q_{ij}>q_{ji}$ then $a_{ij}>0$. The matrix **S** describes the departures from asymmetry, and can be viewed as a matrix with (dis)similarities.

The decomposition has some interesting mathematical and statistical properties. A convenient property is that the matrices **A** and **S** are uncorrelated, the sum of squares of the matrix **Q** can be decomposed into a sum of squares due to symmetry and a sum of squares due to skew-symmetry:

$$\Sigma_i \Sigma_j \, q_{ij}^2 = \Sigma_i \Sigma_j \, s_{ij}^2 + \Sigma_i \Sigma_j \, a_{ij}^2.$$

Because of this split of the sum of squares, the two components can be viewed independently. A related result is that the matrix **S** is the best symmetric approximation to the matrix **Q** in the least squares sense. If we let $\tilde{s}_{ij}$ denote any symmetric quantity, not necessarily the mean of the values in the upper and lower triangle, we always have

$$\Sigma_i \Sigma_j \, (q_{ij} - \tilde{s}_{ij})^2 = \Sigma_i \Sigma_j \, (s_{ij} - \tilde{s}_{ij})^2 + \Sigma_i \Sigma_j \, a_{ij}^2,$$

where the first component of the right hand side is zero by setting $\tilde{s}_{ij} = s_{ij}$.

Under multinomial sampling, the matrix **S** is equal to the maximum likelihood estimator of a symmetric matrix; the hypothesis of symmetry or no important asymmetry can be tested by the chi-square statistic (Bowker, 1948):

$$\chi^2 = \sum_i \sum_{<j} \frac{(q_{ij} - q_{ji})^2}{q_{ij} + q_{ji}}.$$

This statistic follows an approximate chi-square distribution with $n(n-1)/2$ degrees of freedom.

A permutation test for symmetry in a proximity matrix has been developed by Hubert and Baker (1979). The measure is based on the correlation between the elements of a square matrix and its transpose. A test of significance is derived from a reference distribution. Different permutations of the rows and columns yield matrices with different elements in the upper and lower triangles, and hence yield different correlations between the original matrix and the permuted matrix. From these permuted matrices a reference distribution is derived for testing the significance of the observed correlation. A correlation is always between –1 and 1, here perfect positive correlation means symmetry of the proximity matrix and perfect negative correlation means skew-symmetry; a coefficient between these two values indicates that there is some asymmetry in the data.

Transitive Asymmetries

In this section we discuss models that assume that the direction of asymmetry is a *transitive* relation, that is, the following transitivity condition holds for all triples i,j,k: if $q_{ij}>q_{ji}$ and $q_{jk}>q_{kj}$, then $q_{ik}>q_{ki}$. (Holman, 1979; Nosofsky 1991). In terms of skew-symmetry we would have the condition: if $a_{ij}>0$ and $a_{jk}>0$, then we must also have $a_{ik}>0$, for all triples i,j,k.

A general class of models (Holman, 1979), also called *similarity-bias* model (Nosofsky, 1991), gives the proximity of stimulus i to stimulus j by:

$$q_{ij} = F(s_{ij} + r_i + c_j)$$

where F is a general monotonic function, s_{ij} is a symmetric similarity function and r_i and c_j are bias functions on the rows and columns. In the case we choose F equal to the exponential function, the additive model can also be written as a multiplicative model.

We will apply the decomposition to the bias components of the model, $r_i + c_j$, which yields a symmetric similarity and skew-symmetric preference part. Firstly, we define new parameters $u_i = (r_i + c_i)/2$ and $a_i = (r_i - c_i)/2$. The asymmetric part or bias components of the model can now be written as the sum of a skew-symmetric part and a symmetric component

$$r_i + c_j = (u_i + u_j) + (a_i - a_j).$$

The term $u_i + u_j$ is symmetric, and the term $a_i - a_j$ is skew-symmetric, because we always have $a_i - a_j = -(a_j - a_i)$. Here, the bias components have been decomposed into a symmetric part that is usually interpreted as a unicity or specificity effect (Winsberg and Carroll, 1989), and a skew-symmetric part that can be interpreted as a dominance effect. The skew-symmetric part of the models is related to individual properties of the stimuli, that is, skew-symmetry is predicted by the difference function a_i defined on the properties of the individual items. The skew-symmetric part of the bias components is also called the *linear model* for skew-symmetry because the parameters provide a scale for the stimuli, in such a way that skew-symmetry between the stimulus on top and all other stimuli is positive. The difference between two points on this scale is proportional to the magnitude of asymmetry between two stimuli. Thus, in for instance an identification experiment the stimulus on top is more often confused with the other stimuli than that the other stimuli are confused with the stimulus on top. Moreover, if we set the function F of the similarity bias model equal to the identity function we have the stronger condition in terms of skew-symmetry: if $a_{ij} = a_i - a_j$, then

$$a_{ij} + a_{jk} = a_{ik}.$$

The models to be discussed in this section exhibit this form of skew-symmetry or they can be reduced to this form of skew-symmetry after a suitable logarithmic transformation.

The similarity-bias model is a hybrid model (Carroll, 1976; Carroll and Pruzansky, 1981) because it can be viewed as a mixture of two different kinds of models: the similarity function can be continuous and the bias functions can be discrete. The special cases of the similarity bias model are further classified by the structure that restricts the similarity parameters. We have *distance models*, where similarity is related to distance in some psychological space, *similarity-choice models* where similarity is in general unconstrained and *feature models* where similarity is related to a pairwise relation on a feature set.

Distance models

A number of multidimensional scaling models are special cases of this general hybrid model. MDS models that include bias components for the symmetric and skew-symmetric part have been proposed by Weeks and Bentler (1982) and Saito (1991). Explicitly, their model can be written as

$$q_{ij} = k\text{-}F\{d_{ij}(\mathbf{X}) + r_i + c_j\},$$

where $d_{ij}(\mathbf{X})$ is the Euclidean distance function. It is a special case of the similarity-bias model with the Euclidean distance as a dissimilarity function, the function F is set equal to the identity function and r_i and c_j are row and column biases. The constant k, typically chosen as the largest element of $\mathbf{Q}$, converts the dissimilaries given by $F\{d_{ij}(\mathbf{X}) + r_i + c_j\}$ into similarities. Multidimensional scaling models search for a spatial representation of the objects or stimuli in a space of low dimensionality, in such a way that the distances among the n points approximate the dissimilarities as closely as possible. The Euclidean distance function is defined as:

$$d_{ij}(\mathbf{X}) = \sqrt{\Sigma_s \; (x_{is} - x_{js})^2},$$

where x_{is} is the coordinate of object i on dimension s.

The Euclidean distance function satisfies the following axioms:

$d_{ij} \geq d_{ii} = 0$ (minimality)

$d_{ij} = d_{ji}$ (symmetry)

$d_{ij} \leq d_{ik} + d_{jk}$ (triangle inequality)

The minimality axiom states that the distance between two objects should always be greater than or equal to zero; the distance between an object and itself should be zero. The symmetry axiom states that the distance from i to j should be equal to the distance from j to i. The triangle inequality states that the distance from i to j is smaller or equal to the distance from i to j if we travel via k. For overviews of MDS we refer to Kruskal and Wish (1978), Carroll and Arabie (1980), Wish and Carroll (1982) and Coxon (1982).

To illustrate the role of the bias parameters in a multidimensional scaling analysis we apply the model to word association data presented in Table 1. The purpose of the study was to design an advertisement for shampoo. The table entries represent association frequencies with which each column phrase was mentioned when the experimenter presented a row phrase. For further details of the experimental procedure we refer to Harshman, Green, Wind and Lundy (1982).

--- Insert Figure 1 and Table 1 about here---

The word-association data consist of similarity data, and before a distance model can be applied they have to be converted to dissimilarities. This was done by applying a logarithmic transformation to the probabilities q_{ij}, which gives

$$m_{ij} = -\log q_{ij},$$

where the quantities m_{ij} are dissimilarities, which were used as input to the model

$$m_{ij} = d_{ij}(\mathbf{X}) + r_i + c_j.$$

The symmetric part of the biases, which are the averages of the row and columns biases, are displayed on the horizontal axis in the left panel of Figure 1. The skew-symmetric part of the biases are computed by subtracting the row biases from the column biases and dividing the result by two. The skew-symmetric biases are displayed on the vertical axis. Here, we have a positive relationship between the symmetric and skew-symmetric part of the biases. In the Euclidean space of the model, displayed in the right panel of Figure 1, a cluster structure appears consisting of three clusters, the first cluster has members bouncy, manageable and fullness; the second cluster has members natural, not limp, zesty and holds set; the third cluster consists of a single member, body. The phrases in the centre of the configuration have smallest asymmetries and small unicities; large asymmetries and biases are associated with peripheral stimuli.

Others, such as Okada (1988 a,b) proposed a special case of the similarity-bias model with the Euclidean distance function as a similarity function and a linear skew-symmetric function to accommodate asymmetry. The model by Okada has no unicity components, that is, there are no bias components for the symmetric part. This model posits that the (dis)similarity between two objects can be explained by a distance model, and that the bias components are needed for modelling the asymmetry.

The *slide-vector model* attributed to Kruskal in De Leeuw and Heiser (1982), see also Zielman and Heiser (1993), represents asymmetry by adding a vector to the dimensionwise differences. This vector corresponds to a *shift* or *translation* of the points in one direction. The model is written as:

$$q_{ij} = \sqrt{\Sigma_s \, (x_{is} - x_{js} + z_s)^2},$$

where z_s are the elements of the slide vector **z**.

The slidevector model is related to the unfolding model, which associates with each object i a row point $\mathbf{x}_i$ and a column point $\mathbf{y}_i$, with distance

$$d_{ij}(\mathbf{X};\mathbf{Y}) = \sqrt{\Sigma_s\ (x_{is} - y_{js}\ )^2},$$

where only the distances between points of different sets are compared; the distances within each of the two sets are only implicitly defined. The unfolding model considers the symmetric and the asymmetric part of the data as inseparable; see Coombs (1964) or Heiser (1981) for a discussion of the unfolding model. It is not hard to verify that the slide vector model is a special case of the unfolding model with $y_{js} = x_{js} - z_s$.

If we inspect the squared distances, it becomes clear that in the slide-vector model a symmetric and a skew-symmetric part can be distinguished

$$q_{ij}^2 = \Sigma_s\ (\ x_{is} - x_{js}\ )^2 + \Sigma_s\ z_s^2 + 2\Sigma_s\ z_s\ (\ x_{is} - x_{js}\ ).$$

From this decomposition it follows that the model assumes points laying far apart to be more asymmetric than points laying close together on a dimension, because both the symmetric part and the skew-symmetric part are functions of the dimensionwise differences $x_{is} - x_{js}$. The term

$$\Sigma_s z_s\ (\ x_{is} - x_{js}\ )$$

is skew-symmetric, a property that is difficult to recognize in the original form of the model.

---Insert Figure 2 about here---

This representation of asymmetry and dissimilarity is illustrated in Figure 2, where four objects A, B, C, D are depicted as points in a two-dimensional space. The dashed lines in the figure correspond to the projections of the objects on the slide vector. Objects with high projections dominate objects with low projections; in this example object A clearly dominates the other objects. The distances among the points can be interpreted as the similarity or resemblance of the objects; object C is more similar to object D than to object A. The skew-symmetric part of the model is *compensatory*, because the total difference may vanish if large differences on the first dimension are compensated for by differences with opposite sign on the other dimensions. This is the case between objects A and C, because object C projects highest on the second dimension and object A projects highest on the first dimension. The resulting asymmetry is very small, compared to the asymmetry between objects A and B, where A projects highest on both dimensions.

MDS models that do have symmetric bias components but lack the skew-symmetric bias components have been proposed by Bentler and Weeks (1978) and De Leeuw and Heiser (1980). In the context of MDS the symmetric part of the biases can be thought of as unique dimensions (Bentler and Weeks, 1978) or a star-tree (Carroll, 1976); this star-tree can accommodate high centrality or nearest neighbor data (Tversky and Hutchinson, 1986). If object i is a nearest neighbor in the set, the object is the most similar object to all the other objects. Multidimensional scaling imposes a bound on the number of objects that can be the nearest neighbor to an object. However, when the MDS model is extended with additional bias components, this bound is relaxed. If object i is the nearest neighbor of all other objects the corresponding value of the unicity is the smallest among all items.

The distance-density model (Krumhansl, 1978, 1982, 1988) has a similar structure as the previous distance models, but here the bias components are related to the density of points in a region of the multidimensional space. Dissimilarity is modelled as a function of the inter-point distance and the local density of points in the configuration between two objects. The formal structure of the model is:

$$d_{ij}(\mathbf{X},\alpha,\beta) = d_{ij}(\mathbf{X}) + \alpha v_i + \beta v_j,$$

where υ_i is a measure of density of points surrounding point i in the configuration. The distance-density model assumes that within dense subregions of the space finer discriminations are made than within less dense subregions, thus, two points within dense subregions have larger dissimilarities than two points of equal interpoint distance within less dense subregions. The diagonal elements are related to the direction of asymmetry because if i is more similar to j then it is also the case that i has a higher selfsimilarity than j. The symmetric elements of the distance-density model are $d_{ij}(\mathbf{X}) + (\alpha+\beta)(\upsilon_i + \upsilon_j)$, and the skew-symmetric elements are $(\alpha-\beta)(\upsilon_i - \upsilon_j)$

Krumhansl (1978) proposed three measures of density: first, the self similarities (the diagonal of the original table); second, the weighted sum of the distances from an object to the other points, weighted in such a way that the small distances contribute more to the density than the large distances; and thirdly the number of points within a fixed radius of the point. The distance-density model has been extended to tree models by De Sarbo, Manrai and Burke (1990), where density can be estimated from the number of nodes to pass in a hierarchical cluster diagram, or from the number of objects in the same cluster.

Choice models

The similarity choice model (Luce, 1963; Shepard, 1957; Townsend & Landon, 1982), where asymmetry is obtained by multiplying a symmetric term by a factor for rows and a factor for columns, is written as

$$q_{ij} = \frac{\beta_j \, \eta_{ij}}{\Sigma_k \beta_k \, \eta_{ik}}.$$

This is a multiplicative version of the similarity bias model,with F as the exponential function, a similarity function log η_{ij}, a row bias term log β_i and a column bias term $\log\Sigma_k\beta_k \, \eta_{jk}$. The β parameters reflect the tendency to favor some responses over others, and they will account for at least part of the asymmetry in the data. For identification purposes we may require $\Sigma_k\beta_k = 1$ and $\eta_{ii} = 1$ for all i. The η_{ij} parameters are symmetric similarity parameters; these parameters can, after transformation to dissimilarity, be further analyzed by a MDS program. The transformation from similarities to dissimilarities is often done by a logarithmic function (Heiser, 1988). Note that the normalizing term in the denominator, can also be seen as a row-specific factor accounting for asymmetry, since even if $\beta_j = 1$ for all j, **Q** will generally not be symmetric.

The quasi-symmetry model (Caussinus, 1965) is very similar to the similarity choice models and also supposes a symmetric model for pairs, and parameters for the column categories β_j, but here the parameters for the row categories α_i cannot be expressed in terms of the other parameters. Explicitly, the model is written as

$$q_{ij} = \alpha_i \, \beta_j \, \eta_{ij}.$$

To identify the parameters of the model, we have to specify some constraints, a discussion of possible constraints can be found in Constantine and Gower (1982).

When the row sums are different, we usually apply the quasi-symmetry model. When the rows sum to one, the parameters of the quasi-symmetry model can be expressed in the form of the similarity choice model. In the similarity choice model, the α_i parameters appear in the denominator which is expressed in terms of the η_{ij} and β_j parameters. Reasoning along similar lines as in the choice model, we may rewrite the quasi-symmetry model as a similarity bias model with F an exponential function and $s_{ij} = \log \eta_{ij}$, $r_i = \log \alpha_i$ and $c_j = \log \beta_j$.

A least squares version of the quasi-symmetry model has been proposed by Levin and Brown (1979). Their method maximizes symmetry of the rescaled datamatrix using a least squares criterion. The analysis yields rescaling coefficients, which can be interpreted in a similar way as the quasi-symmetry model. The procedure can be described as trying to fit a model of the form

$$q_{ij} v_i = s_{ij},$$

where v_i are rescaling coefficients. When a perfect rescaling is possible in the sense that **S** is indeed symmetric, we may rewrite this equation as $q_{ij} = \frac{s_{ij}}{v_i}$. The quasi-symmetry model can be written in this form by defining $\tilde{\alpha}_i = \frac{\alpha_i}{\beta_i}$ and $\tilde{\eta}_{ij} = \beta_i \beta_j \eta_{ij}$. Then the model can be written as

$$q_{ij} = \tilde{\alpha}_i \tilde{\eta}_{ij}.$$

Thus, in the case we have a perfect fit, Levin and Brown's (1979) procedure becomes a way to fit the quasi-symmetry model with $\tilde{\alpha} = \frac{1}{v_i}$. Applying the decomposition into a symmetric and skew-symmetric component we have for the skew-symmetric elements:

$$a_{ij} = (\tilde{\alpha}_i - \tilde{\alpha}_j)\, \tilde{\eta}_{ij}.$$

This form of skew-symmetry differs from the linear model because we have applied the decomposition to the untransformed elements instead of the logarithms of the data elements.

A related possibility for dealing with asymmetry is to transform the asymmetry away by monotone or some other form of regression. This is usually done simultaneously with a multidimensional scaling program (Kruskal, 1964 a,b). This can be done row-conditionally which means that values *within* rows are regarded as comparable with each other, and values *among* the rows are regarded as incomparable. In the case of linear transformations with no additive constant in the regression equation, this approach is basically the same as the procedure of Levin and Brown (1979), because the target (the distance matrix) is symmetric, so that the rescaling will optimize the symmetry of the transformed data as well. If the data are linearly transformed, the regression weights can be interpreted as bias parameters or the tendency to favour some responses over others.

Feature models

Tversky (1977) challenged the dimensional-metric assumptions that underlie the geometrical approach to the analysis of similarity. From a set theoretical viewpoint the feature matching model was developed. The feature matching model assumes that each object is characterized by a set of features. The similarity between objects i and j is expressed as a function of their common and distinctive features. The additive version of the model is called the contrast model. In terms of similarities we have:

$$q_{ij} = \theta \, \mathrm{f}\,(i \cap j) - \alpha \, \mathrm{f}\,(i - j) - \beta \, \mathrm{f}(j - i),$$

where $(i \cap j)$ is the number of features shared by objects i and j, $(i - j)$ are the set of features unique to object i with respect to object j, $(j - i)$ is the set of unique features belonging to object j with respect to object i and f is a measure function of the features. The psychological content of the model is discussed in Tversky and Gati (1978). The parameters θ, α, β are assumed to be positive; they must be estimated from the data, which can be done by linear regression techniques. The function f measures the contribution of the individual features to the similarity between the objects. If the values of α and β are different the model represents asymmetry; if the values are equal we have a symmetric model. In addition to representing asymmetry, the model describes differences in self-similarities of the objects as well. The feature matching model implies that if i is more similar to j it must be true that j is more self similar than i (Nosofsky, 1991).

The additive version of the feature matching model can be conveniently decomposed into a symmetric and skew-symmetric part, where the symmetric part consists of a common features term and a term representing the number of features, the skew-symmetric part of the data is predicted solely from the number of features. First we note that the feature matching model can be rewritten as

$$q_{ij} = (\theta + \alpha + \beta)\mathrm{f}\,(i \cap j) - \alpha \, \mathrm{f}\,(i) - \beta \, \mathrm{f}(j),$$

where the asymmetric biases represent the number of features. The symmetric part is given by

$$s_{ij} = \frac{1}{2}(\theta + \alpha + \beta)\mathrm{f}\,(i \cap j) + (-\alpha - \beta)\,\mathrm{f}\,(i) + (-\alpha - \beta)\beta\,\mathrm{f}(j),$$

and the skew-symmetric part is given by

$$a_{ij} = \frac{1}{2}(\beta - \alpha)\, f(i) - (\beta - \alpha) f(j),$$

The major problem of the feature matching model is that application of the model is limited to the case where a suitable feature set can be defined. This problem is sometimes relatively easy to solve, for instance in case of a letter confusion matrix (Keeren and Baggen, 1981), and sometimes difficult to solve, for instance in the case of a matrix with sociometric interactions.

Additive trees (Sattath and Tversky, 1977) represent objects as "leaves" on a tree in such a way that distances calculated between the leaves via the arcs of the tree correspond as closely as possible to the dissimilarities. Additive trees are also known under the name "free" tree (Cunningham, 1978) and "path length tree" (Carroll, 1976). As explained by Sattath and Tversky, a tree is a distinctive feature model, where each arc represents a feature, and stimuli that follow from that arc possess a feature. The tree distance between two stimuli can be written as

$$d_{ij} = \Sigma_k \alpha_k \,(1\text{-}p_{ik}) p_{jk} + \Sigma_k \alpha_k \,(1\text{-}p_{jk}) p_{ik} \,,$$

where α_k is a measure of feature k, and p_{ik} denotes an indicator variable; if the value of the indicator variable is one, object i possesses feature k, and if the indicator variable is zero, feature k is not a feature of object i. Cunningham (1978) generalized the tree to a bidirectional tree by differentially weighting the path lengths

$$d_{ij} = \Sigma_k \alpha_k \,(1\text{-}p_{ik}) p_{jk} + \Sigma_k \beta_k \,(1\text{-}p_{jk}) p_{ik} \,,$$

thus if the subscripts of d_{ij} are reversed, the set of weights differ. Applying the decomposition, we have for the symmetric part

$$s_{ij} = \frac{1}{2}\Sigma_k(\alpha_k + \beta_k)\{(1-p_{ik})p_{jk} + (1-p_{jk})p_{ik}\},$$

which can be interpreted as an additive tree with averaged weights. The skew-symmetric part of a bidirectional tree can be written as

$$a_{ij} = \frac{1}{2}\Sigma_k(\alpha_k - \beta_k)\{(1-p_{ik})p_{jk} - (1-p_{jk})p_{ik}\},$$

which can be simplified further by expanding the product terms as

$$a_{ij} = \frac{1}{2}\Sigma_k(\alpha_k - \beta_k)\{p_{jk} - p_{ik}\}$$
$$= a_j - a_i.$$

with a_i defined as $\frac{1}{2}\Sigma_k(\alpha_k - \beta_k)p_{ik}/2$. The present analysis shows that a bidirectional tree can be written as the sum of a symmetric additive tree and skew-symmetric biases for modelling asymmetry. A useful result for data analysis is that we may construct a bidirectional tree from an additive tree from the symmetric part and a linear model from the skew-symmetric part. It is well known that the least squares estimates of the linear model can be obtained by computing the row means from a skew-symmetric matrix. The last step in this methodology is to stretch or shrink the unique arcs of the tree. This can be done by doubling these arcs and then adding the row-mean of object i to one arc and subtracting the row mean from the other.

Dobson (1974) has shown that for symmetric proximities an additive tree can be decomposed into an ultrametric tree and a star tree. A star tree is a tree with one internal node connecting all objects and this structure can also be represented by the linear form $u_i + u_j$. This result can also be applied to bidirectional trees, because it has been shown that the symmetric part is represented by an additive tree. The implication is that a bidirectional tree is also represented as a similarity-bias model with asymmetric biases and an hierarchical tree representation for the similarity part. We conclude that a bidirectional tree can be represented as a similarity bias model with either an additive tree with skew-symmetric biases or an hierarchical tree with asymmetric biases.

Intransitive or circular asymmetries

In the previous section we have discussed models that assume that asymmetry is a transitive relation, the present section discusses models that relax this condition. The transitivity condition need not hold for all triples i,j,k, that is, we may still have transitivity: if $q_{ij}>q_{ji}$ and $q_{jk}>q_{kj}$, then $q_{ik}>q_{ki}$. but we may also have $q_{ik}<q_{ki}$. In terms of skew-symmetry we would recognize an intransitive triple by: if $a_{ij}>0$ and $a_{jk}>0$, then we have $a_{ik}<0$, for some triples i,j,k. In the choice literature intransitivities are also called circular triads, because when three stimuli are represented in a graph, the directed arrows indicating the sign of the relation follow a circle.

Decomposition of a skew-symmetric matrix

Gower (1977), Constantine and Gower (1978) and Gower and Digby (1981) studied the singular value decomposition of the skew-symmetric matrix **A**. The singular value decomposition is a bilinear method, which means that there are two sets of parameters, each of which forms a linear function with respect to the other. It decomposes any matrix **B** into a product of the form:

$$\mathbf{B} = \mathbf{W} \Lambda \mathbf{V}'.$$

Here **W** and **V** are both orthogonal matrices, i.e. **W'W** = **I** and **V'V** = **I**, and Λ is a diagonal matrix with singular values. For a skew-symmetric matrix **A** the singular values come in pairs, i.e. Λ contains the singular values λ_1, λ_1,..,$\lambda_{n/2}$, $\lambda_{n/2}$, with the last singular value being equal to zero when n is odd. Due to this peculiarity, the singular value decomposition of a skew-symmetric matrix can be rewritten into a form that better expresses its fundamental structure:

$$\mathbf{A} = \mathbf{W}\,\Lambda\,\mathbf{J}\,\mathbf{W}',$$

where **W** and **J** are again orthogonal matrices and Λ is the diagonal matrix of singular values as defined above. The matrix **J** is a block diagonal matrix with 2 by 2 sub-matrices with zero's on the diagonal, 1 above the diagonal and -1 below the diagonal. When n is odd the last diagonal position is filled with a zero. The presence of **J** makes the left singular vectors **W** a permutation and reflection of the right singular vectors **WJ**'. A one-dimensional representation of skew-symmetry with the SVD does not exist (Gower, 1977).

If we denote the first column of **W** by **e** and the second by **f**, the two-dimensional model is given by the elements

$$a_{ij} = \lambda_1(e_i f_j - f_i e_j). \tag{2}$$

If we define $g_i = \frac{e_i}{f_i}$ we may also write the equation above as the product of a linear part and skew-symmetry part

$$a_{ij} = \lambda_1 f_i f_j\,(g_i - g_j).$$

A necessary condition for occurence of intransitivities is that some of the products $f_i f_j$ are negative while others are positive; however, this is not a sufficient condition. A small example with three objects is given by

$\mathbf{f}=\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ and $\mathbf{g}=\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, which yields the predicted values

$$\begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix}.$$

Here, object 1 is preferred over object 2, and object 2 is preferred over object 3, but it is not true that object 1 is preferred over object 3.

Interpretation of skew-symmetry analysis is simplified by plotting the objects o_i with coordinates (e_i,f_i) in a two dimensional space. In such a diagram, twice the area of the triangle with vertices at the two points and the origin O is an approximation of the element a_{ij}. This diagram is also called a Gower diagram. The areas of the triangles O o_i o_j and O o_j o_i are equal, but they have opposite sign, thus modelling skew-symmetry. Two points may be far apart while there is still a perfect symmetric relation, this happens when two points are located on a line that passes through the origin. The representation of points by vectors will be more useful; this is illustrated in Figure 3.

---Insert Figure 3 about here---

Three objects are depicted in Figure 3; the relation between objects *j* and *k* is symmetric, the relations of these two objects to object *i* are asymmetric. For two vectors of equal length, the greater the angle between two vectors, the larger the asymmetry.

If all points are collinear on a line (or almost collinear) the asymmetric part of the data can be modeled by a linear skew-symmetric function. This special case has been studied by Weeks and Bentler (1982), Okada (1988 a,b), Takane and Shibayama (1986) and Holman (1979) and – in the context of Thurstone case V scaling – by Mosteller (1951) and Gulliksen (1956). This simple linear form of skew-symmetry can be written as:

$$a_{ij} = a_i - a_j.$$

Note that this equation is obtained from equation (2) if we substitute $e_i = e_j = 1$ and $a_i = \lambda_1 f_i$.

We will now illustrate the SVD of a skew-symmetric matrix with an example.

The Rothkopf (1957) Morse Code Data

Rothkopf (1957) collected a set of confusions among 36 auditory Morse code signals. The signals, given in Table 2, are composed of a series of short and long signals. Subjects listened to two signals and were required to state whether the two signals presented were the same or different. Table 2 presents a portion of this larger table. Each number in the table is the proportion of 150 respondents who responded "same" to the row signal followed by the column signal. The singular values are .309, .115, .114, .051 and .018. Because the singular values come in pairs, a model that consists of one Gower diagram would account for 76 percent of the variance in the skew-symmetric part of the table. The first Gower diagram, shown in Figure 4, shows asymmetries that occur due to changes between long tones and short tones. In the lower right corner of the diagram we find digits where the switch between different lengths in the signal occurs after the first or second tone (1,2,6,7); in the upper right corner we find the switch after the third and fourth tones (3,4,8,9); in the upper left corner we find the homogeneous tones (5,0). The largest skew-symmetry in the diagram is between digit 5 and 6.

---Figure 4 and Table 2 about here---

Small asymmetries occur between dissimilar pairs such as (1,6) and (2,7), because these pairs are close together on the map.

The DEDICOM model

The DEDICOM (DEcomposition into DIrectional COMponents) model was proposed by Harshman, Green, Wind and Lundy (1982), and decomposes an asymmetric matrix as:

$$\mathbf{Q} = \mathbf{L}\ \mathbf{R}\ \mathbf{L}'$$

The matrix **L** denotes an n by p matrix of loadings of the n observed objects onto p dimensions or aspects of the objects, and the matrix **R** of order p by p is an *asymmetric* matrix describing the directional relationships among the components or dimensions. Thus, according to the DEDICOM model any asymmetry between the objects is due to asymmetric relations between the dimensions. The DEDICOM model is a special case of factor analysis or components analysis in the sense that the factor loadings for the rows are a linear transformation of the factor loadings of the columns. It should be noted that the one-dimensional model is symmetric.

Applying the general tactic again, the matrix **R** can be decomposed into a symmetric part **C** and a skew-symmetric part **T**. As a result of the distributive properties of matrix multiplication the model can be decomposed into a symmetric part and a skew-symmetric part:

$$\mathbf{Q} = \mathbf{L}\ \mathbf{R}\ \mathbf{L}' = \mathbf{L}\ \mathbf{C}\ \mathbf{L}' + \mathbf{L}\ \mathbf{T}\ \mathbf{L}'.$$

It follows that the DEDICOM model is an additive scalar product model; the model is the sum of an oblique factor model for the similarity part and a skew-symmetric function of the factors for the dominance part. Because the structure of the skew-symmetric part of DEDICOM is equivalent to a SVD of a skew-symmetric matrix, DEDICOM also accomodates the possibility of circular triads.

There is a rotational indeterminacy in the model. This rotational indeterminacy may be used to rotate the matrix with loadings to obtain a simple diagram. The matrix **L** may be rotated if we pre- and postmultiply **R** by the inverse of the chosen rotation matrix. A convenient choice of a rotation matrix is the matrix with singular vectors of the SVD of the matrix **T**. If the matrix **L** is rotated by this rotation matrix the diagram of two dimensions can be interpreted as follows: the angle of two vectors multiplied with the obliqueness of the axis corresponds to the symmetric part of the matrix **Q**, and the area of the triangle corresponds to the skew-symmetric part of the model as in the SVD method of a skew-symmetric matrix. If we choose this rotation matrix, the matrix **R** is symmetric except in its 2 by 2 diagonal blocks. The first dimension is asymmetrically related to the second dimension and not to any other dimension. Of course the SVD of the matrix **T** remains interesting in its own right because a plot of the singular vectors shows graphically how the dimensions are related.

If we have a two-dimensional solution there is another useful rotation. This rotation method shows the close relationship with the SVD. First, we compute the eigen decomposition of the matrix $\mathbf{C} = \mathbf{K}\Lambda\mathbf{K}'$. Second, assuming the inverse $\Lambda^{-1/2}$ exists, the matrix **L** is rotated with the scaled eigen vectors $\mathbf{K}\Lambda^{-1/2}$ of the symmetric matrix **C**. The matrix **T** is a 2 by 2 skew-symmetric matrix that can be written as $\tau\mathbf{J}$, where τ is a scalar and **J** $= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The matrix **T** is rescaled by this rotation with values $\lambda_1\lambda_2$. Using this rotation the matrix **C** reduces to the identity matrix. When the objects are plotted using the coordinates of $\mathbf{LK}\Lambda^{-1/2}$, the area between the points and the origin corresponds to the skew-symmetric part, while the angle between points corresponds to the symmetric part.

The two-dimensional DEDICOM model has been proposed under the name ASYMSCAL by Chino (1978). The **R** matrix in Chino's method has a very special structure:

$$\mathbf{R} = \alpha\mathbf{I} + \beta\mathbf{J},$$

where α,β are parameters to be estimated, **I** is the identity matrix of order 2 by 2 and **J** is a 2 by 2 skew-symmetric matrix with a one above the diagonal; the cell below the diagonal is filled with -1. After the rotation of the two-dimensional DEDICOM described previously, the **R** matrix of the DEDICOM model has similar structure with α=1. The models are identical because we may rescale the matrix with loadings, if we adjust the **R** matrix by the inverse of the rescaling factor.

Algorithms for fitting the DEDICOM model are given by Kiers (1989) and Kiers et al.(1990). An algorithm for fitting the off-diagonal DEDICOM model is given by Ten Berge and Kiers (1989). This algorithm is nearly identical to Chino's (1978) algorithm; the difference between these two algorithms is that the **R** matrix is estimated in different ways.

ASYMSCAL

Young (1975, 1984, 1987) proposed a weighted Euclidean distance model to represent asymmetry:

$$d_{ij}^2(\mathbf{X}) = \Sigma_s w_{is} \, (x_{is} - x_{js})^2,$$

where the weights w_{is} are specific for each stimulus i and dimension s. When the weights are unity the model reduces to the Euclidean distance model. The asymmetry is accounted for by row or stimulus weights operating on the dimensions of the configuration. For each stimulus the dimensions of the configuration must be adjusted by shrinking or stretching the axes. Thus, for every stimulus the analysis yields p row weights resulting in n by p additional parameters, and also n different configurations. A remedy to reduce the number of parameters is to average the weights over meaningful clusters of objects and then plot spaces for groups of objects using the averaged weights.

The ASYMSCAL procedure has skew-symmetric elements that is the product of a linear function and a symmetric squared distance component:

$$\tfrac{1}{2}\,\Sigma_s(w_{is} - w_{js})(x_{is} - x_{js})^2.$$

The one dimensional version of the model satisfies the transitivity condition, because the distances are positive. In one dimension the skew-symmetric elements are similar to those of the choice models. In this case the model can be interpreted as a similarity bias model with F an exponential function, log $(x_{is} - x_{js})^2$ a similarity function, $\log w_{is}$ a bias term for rows and no bias terms for columns.

In more than one dimension, the model can represent intransitivities because different contributions of the skew-symmetric part have to be added. The differences in skew-symmetry between the object weights are larger for dissimilar objects, because these differences are weighted by $(x_{is} - x_{js})^2$.

Discussion

Our point of view that asymmetric proximities are a combination of similarity and dominance provides a classification of methods for asymmetric proximities. The first class of methods that have been described, usually called the similarity bias model, predicts asymmetry from a linear function. In this paper it is shown that additional models are special cases of this general model (Weeks and Bentler, 1982; Okada, 1988a, 1988b; Saito, 1991; Levin and Brown, 1979; Cunningham, 1978). Because of this linearity, the asymmetries predicted by the model satisfy the transitivity condition. The second class of models have the possibility of representing intransitivities or circular triads. Another difference between the two classes is that the skew-symmetric part of the models in the intransitive class are multidimensional, whereas the skew-symmetric part of the transitive models are one-dimensional. The one-dimensional versions of the models from the first class are not interesting or do not exist.

The decomposition of bias parameters into a dominance and a specificity component provides a more natural interpretation than row and column biases. These interpretations of specificity and dominance are generic names, for the symmetric and skew-symmetric parts of the biases. In applications, the process or processes to which these components refer can hopefully be identified. The feature matching model and the distance density model deserve special attention here, because two components refer to the number of features and the density of points, respectively. The slide-vector model and the DEDICOM model provide examples where dominance is a function of the similarity parameters.

Although a number of models have been proposed for the analysis of asymmetric data tables, software for computing the parameters is not readily available. Packages like SPSS, BMDP or SAS only offer the model of quasi-symmetry and the similarity-choice model. ASYMSCAL is an option in the ALSCAL program, which is available in the SPSS program. Even the SVD of a skew-symmetric matrix is difficult to obtain. Although most programs have a procedure for principal components analysis, this procedure works after standardizing the variables. For the other models researchers have to write the programs themselves, or ask for prototypes.

Progress in this area can be expected to come from software development that is easily accessible. For the models satisfying the transitivity condition, we may see three-way models, where asymmetry is studied in different conditions in the future. For the models that may represent intransitive asymmetries, it seems less clear where progress can be made.

References

Bentler, P. M. & Weeks, D. G. (1978). Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, *17*, 138-151.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, *43*, 572-574.

Carroll, J. D. (1976). Spatial, non-spatial and hybrid models for scaling. *Psychometrika*, *41*, 439-463.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In: M.R. Rosenzweig and L. W. Porter (Eds), *Annual Review of Psychology*, *31*, 607-649.

Carroll, J. D., & Pruzansky, S. (1981). Discrete and hybrid scaling models. In E.D. Lantermann & H. Feger (Eds), *Similarity and Choice*. Huber, Bern, 108-139.

Caussinus, H. (1965). Contribution a l'analyse statistique des tableaux de correlation. Annals of the faculty of science, University of Toulouse, 29, 77-182.

Chino, N. (1978). A graphical technique for representing the asymmetric relationships between *N* objects. *Behaviormetrika*, *5*, 23-40.

Clogg, C. C. (1981). Latent structure models of mobility. *American Journal of Sociology*, *86*, 836-868.

Constantine, A. G., & Gower, J. C. (1978). Graphical representation of asymmetric matrices. *Journal of the Royal Statistical Society(series C)*, *27*, 297-304.

Constantine, A. G., & Gower, J. C. (1982). Models for the analysis of interregional migration. *Environment and Planning A*, *14*, 477-497.

Coombs, C. H. (1964). *A theory of data*. Wiley, New York.

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling*. Heinemann Educational Books LTD, London.

Cunningham, J. P. (1978). Freetrees and bidirectional trees as representations of psychological distance. *Journal of Mathematical Psychology*, *17*, 165-188.

De Leeuw, J., & Heiser, W.(1980). Multidimensional scaling with restrictions on the configuration. In: P. R. Krisnaiah (Ed.), *Multivariate analysis-V.* North Holland, Amsterdam, 501-522.

De Leeuw, J., & Heiser, W. (1982). Theory of multidimensional scaling. In P. R. Krisnaiah and L. N. Kanal (Eds.), *Handbook of statistics*,Vol 2. North Holland, Amsterdam, 285-316

DeSarbo, W. S., Manrai, A. K., & Burke, R. R. (1990). A non spatial methodology for the analysis of two way proximity data incorporating the distance density hypothesis. *Psychometrika, 55*, 229-253.

DeSarbo, W. S. & De Soete, G. (1984). On the use of hierarchical clustering for the analysis of nonsymmetric proximities. *Journal of Consumer Research*, 11, 601-610.

Dobson, A. J. (1974). Unrooted trees for numerical taxonomy. *Journal of Applied Probability*, *11*,32-42.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In: J. R. Barra, F. Brodeau, G. Romer, & B. van Cutsem (Eds.), *Recent developments in statistics.* North Holland, Amsterdam, 109-123.

Gower, J. C., & Digby, P. G. (1981). Expressing complex relationships in two dimensions. In: V. Barnett (Ed), *Interpreting multivariate data.* Wiley, New York, 83-118.

Gulliksen, H. (1956). A least squares solution for paired comparisons with incomplete data. *Psychometrika*, *20*, 125-134.

Harshman, R. A., Green, P. E., Wind, Y., & Lundy, M. E., (1982). A model for the analysis of asymmetric data in marketing research. *Marketing Science, 1*, 204-242.

Harshman, R. A., & Lundy, M. E. (1990). Multidimensional analysis of preference structures. In: A. de Fontenay, M. H. Shugard, & D. S., Sibley (Eds) Telecommunications demand modelling: An integrated view (pp. 185-204). Amsterdam: Elsevier.

Heiser, W. J. (1981). Unfolding analysis of proximity data. Doctoral thesis, Department of Data Theory, University of Leiden, The Netherlands.

Heiser, W. J. (1988). Selecting a stimulus set with prescribed structure from empirical confusion matrices. *British Journal of Mathematical and Statistical Psychology. 41*, 37-53.

Holman, E. W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology, 20*, 1-15.

Hubert, L. J., & Baker, F. B. (1979). Evaluating the symmetry of a proximity matrix. *Quality and Quantity*, *43*, 81-91.

Keeren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception and psychophysics*, *29*, 234-246.

Kiers, H. A.L. (1989). An alternating least squares algorithm for fitting the two and three way DEDICOM model and the IDIOSCAL model. *Psychometrika*, *54*, 515-521.

Kiers, H. A. L., Ten Berge, J. M. F., Takane, Y., & De Leeuw, J. (1990). A generalization of Takane's algorithm for DEDICOM. *Psychometrika*, *55*, 151-159.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *84*, 445-463.

Krumhansl, C. L. (1982). Density versus feature weights as predictors of visual identifications. Comments on Appelman and Mayzner. *Journal of experimental psychology: general*, *111*, 101-108.

Krumhansl, C. L. (1988). Testing the density hypothesis: comment on Corter. *Journal of experimental psychology: general*, *117*, 101-104.

Kruskal, J. B. (1964 a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* , **29**, 115-129.

Kruskal, J. B. (1964 b).Nonmetric multidimensional scaling: a numerical method. *Psychometrika* , **29**, 1-28.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. *Sage university paper series on quantitative applications in the social sciences*, 07-011. Beverly Hills and London: Sage Publications.

Levin, J., & Brown, M. (1979). Scaling a conditional proximity matrix to symmetry. *Psychometrika*, *44*, 239-243.

Luce, R. D. (1963). Detection and recognition. In: R.D. Luce et al (Eds.), *Handbook of Mathematical Psychology.* Vol I. Wiley, New York.

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, *16*, 3-9.

Noble, B. & Daniel, W. (1988). *Applied linear algebra.* Prentice-Hall, New Jersey.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*, 94-140.

Okada, A. (1988 a). Asymmetric multidimensional scaling of car switching data. In: W. Gaul & M. Schader (Eds.), *Data, expertknowledge and decisions.* Springer verlag, Berlin, 279-290.

Okada, A. (1988 b). An analysis of intergenerational occupational mobility by asymmetric multidimensional scaling. *Proceedings of the SMABS 88 conference.* University of Groningen, 1-15.

Rao, V. R. & Sabavalla, D. J. (1981). Inferences from hierarchical choice processes from panel data, *Journal of consumer research*, *8*, 85-96.

Rosch, E. H. (1975). Cognitive reference points. *Cognitive Psychology*, *1*, 532-547.

Rothkopf, E. Z. (1957) A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 53, 94-101.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, *42*, 319.

Saito, T. (1986). Multidimensional scaling to explore complex aspects in dissimilarity judgment. *Behaviormetrika*, *20*, 35-62.

Saito, T. (1991). Analysis of asymmetric proximity matrix by a model of distance and additive terms. *Behaviormetrika*, *29*, 45-60.

Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika, 22*, 325-345.

Ten Berge, J. M. F., & Kiers, H. A. L. (1989). Fitting the off-diagonal DEDICOM model in the least-squares sense by a generalization of the Harman and Jones minres procedure of factor analysis. *Psychometrika*, *54*, 333-337.

Takane, Y., & Shibayama, T. (1986). Comparison of models for stimulus recognition data. In: J. De Leeuw e.a. (Eds), *Multidimensional data analysis*, DSWO press, Leiden, 119 - 134.

Tobler, W. (1976). Spatial interaction patterns. *Journal of Environmental Systems*, 1976, 6, 271-301.

Townsend, J. T., & Landon, D. E. (1982). An experimental and theoretical investigation of the constant-ratio rule and other models of visual letter confusion. *Journal of Mathematical Psychology*, *25*, 119-162.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-252.

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & S. B. B. Lloyd (Eds), *Cognition and categorization.* Lawrence Erlbaum associates. Hillsdale, New Jersey, 79-98.

Tversky, A., & Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review*, *93*, 3-22.

Urban, G. L., Johnson, P. L. & Hauser, J. R. (1984). Testing competative market structrures. *Marketing Science*, *3*, 83-112.

Weeks, D. G., & Bentler, P. M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika*, *47*, 201-208.

Winsberg, S., & Carroll, J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika*, 54,217-229.

Wish, M., & Carroll, J. D. (1982). Multdimensional scaling and its applications. In P. R. Krisnaiah and L. N. Kanal (Eds.), *Handbook of statistics,* Vol 2. North Holland, Amsterdam, 317 - 345.

Young, F. W. (1975). An asymmetric Euclidean model for multiprocess asymmetric data. Paper presented at the US-Japan seminar on multidimensional scaling, University of California, San Diego, La Jolla.

Young, F. W. (1984). The general Euclidean model. In: H. G. Law et al. (Eds.), *Research methods for multimode data analysis.* Praeger, New York, 440-470.

Young, F. W. (1987). *Multidimensional scaling*. Lawrence Erlbaum associates. Hillsdale, New Jersey.

Zielman, B., & Heiser, W. J. (1993). Analysis of asymmetry by a slide-vector. *Psychometrika*, *58*,101-114.

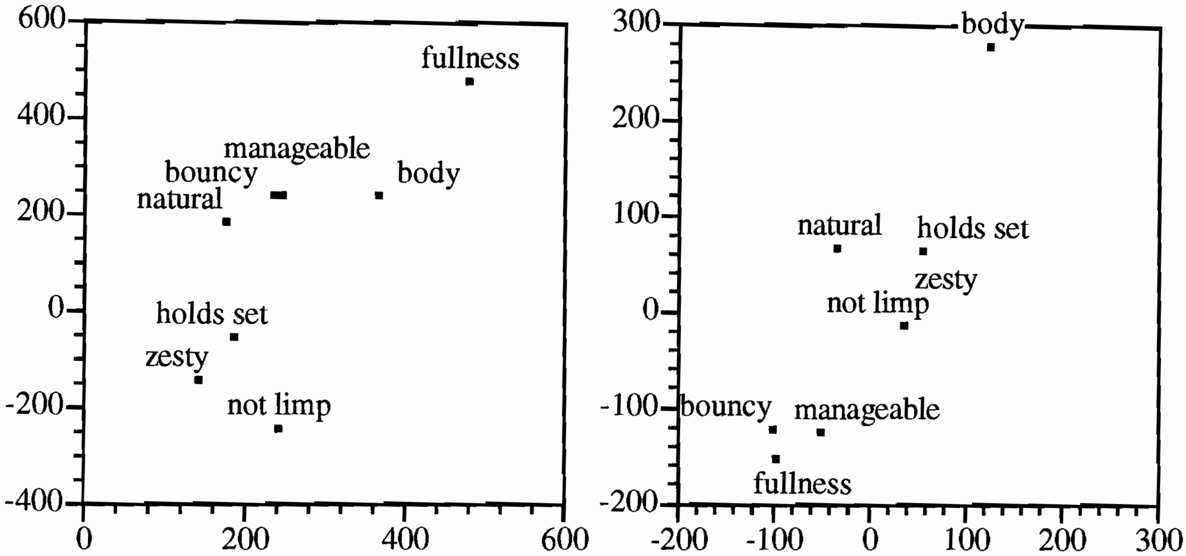

Figure 1: Left panel: on the horizontal axis the symmetric part of the bias components, on the vertical axis the skew-symmetric part of the bias components; right panel: the scaling configuration.

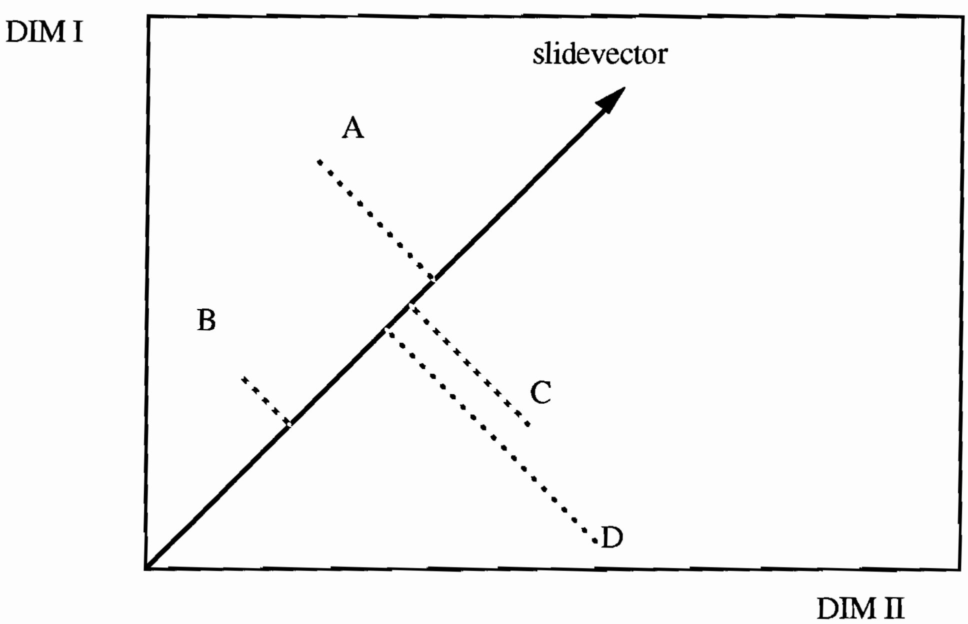

Figure 2: Joint representation of the symmetry and asymmetry by the slide-vector model

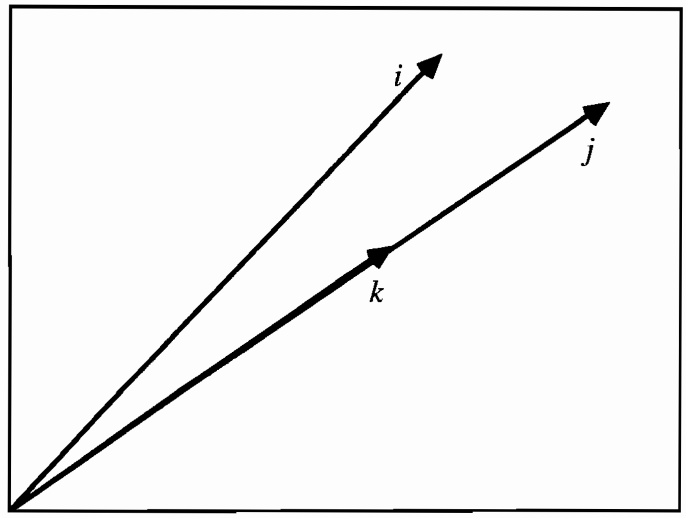

Figure 3: Representation of asymmetry by the bilinear model

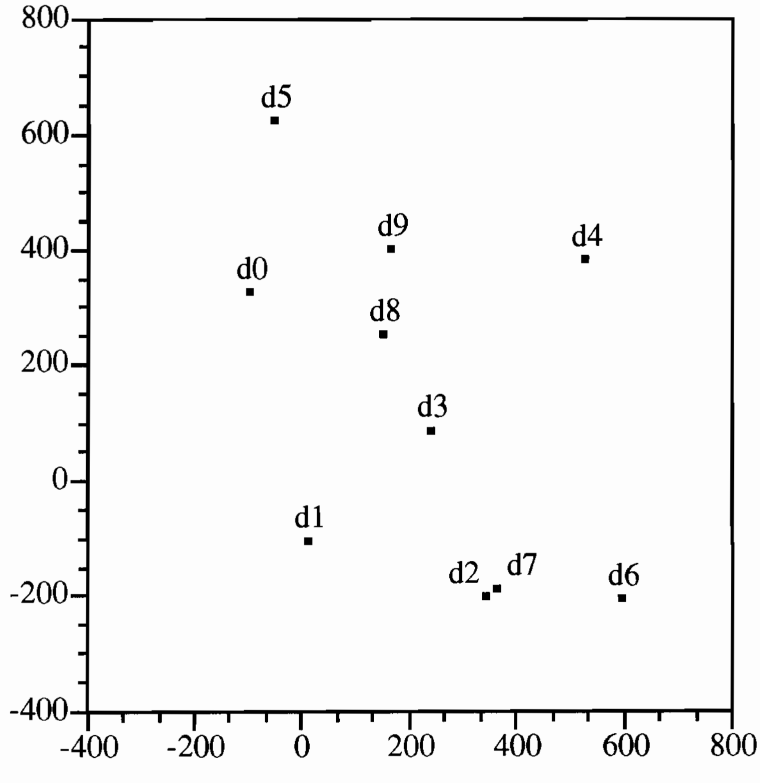

Figure 4: First Gower diagram for the Morsecode data

Table 1: Word association data; rows refer to words presented and columns refer to responses

Body	-	44	5	23	1	19	1	3
Fullness	22	-	5	3	1	9	1	2
Holds Set	17	21	-	5	0	17	0	5
Bouncy	15	12	3	-	1	5	0	14
Not Limp	28	27	4	18	-	4	1	7
Manageable	17	13	11	2	0	-	0	3
Zesty	7	9	2	22	0	4	-	13
Natural	4	9	1	2	0	7	1	-

Table 2: Digits in the Morse code confusions data

Stimulus	Signal										
1	.----	84	63	13	08	10	08	19	32	57	55
2	..---	62	89	54	20	05	14	20	21	16	11
3	...--	18	64	86	31	23	41	16	17	08	10
4	....-	05	26	44	89	42	44	32	10	03	03
5	.....	14	10	30	69	90	42	24	10	06	05
6	-....	15	14	26	24	17	88	69	14	05	14
7	--...	22	29	18	15	12	61	85	70	20	13
8	---..	42	29	16	16	09	30	60	89	61	26
9	----.	57	39	09	12	04	11	42	56	91	78
0	-----	50	26	09	11	05	22	17	52	81	94